JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 39, 1305-1319 (2023)
DOI: 10.6688/JISE.202311_39(6).0004

A Schizophrenia Screening Method Based on Fine-grained Voice Features and Network Isolation Privacy Protection

XIANCHUAN LING[1], XIN ZHANG[1,+], HUIRU WANG[1], GUOZHU ZHAO[2], HUANHUAN ZHAO[2], SHENGHUI ZHAO[2] AND ZENGHUI DING[3]

[1]*Chuzhou Second People's Hospital, Anhui, 239000 P.R. China*
E-mail: zhangxin0550@126.com
[2]*Chuzhou University, Anhui, 239000 P.R. China*
E-mail: gzzhao2015@126.com
[3]*Institute of Intelligent Machines, Hefei Institutes of Physical Science*
Chinese Academy of Sciences, Anhui, 230031 P.R. China
E-mail: dingzenghui@iim.ac.cn

Schizophrenia screening and risk assessment is critical to social harmony and stability as well as personal health and growth. Based on audio data from 179 schizophrenia convalescent patients, this paper proposes a Support Vector Machine (SVM) based schizophrenia screening and risk assessment framework utilizing network physical isolation within a private cloud for user privacy protection. We first design a network security strategy based on wireless sensor network using network physical isolation technology within a private cloud to ensure the absolute security of users' personal privacy. We further exploit the one-class classification technique to formulate the schizophrenia screening decision process as an SVM model. Extensive experiments are conducted to illustrate the schizophrenia screening performance of the proposed SVM based screening framework in terms of accuracy. We also investigate the related authentication efficiency issues in terms of the usability to audio contents and the scalability to the number of audio features.

Keywords: schizophrenia screening, wireless sensor networks, SVM, network physical isolation, privacy protection

1. INTRODUCTION

Schizophrenia is a serious mental disorder, which can lead patients to interpret reality in an abnormal way. Schizophrenia also may lead to hallucinations, delusions and some extreme confusion of thinking and behavior, affect the daily life of patients, and cause disability. Extensive research shows that early detection and screening of schizophrenia patients can effectively control and manage schizophrenia, and prevent the adverse effects of schizophrenia patients on society, family and themselves. Notice that by now there is still a lack of effective early schizophrenia screening methods and techniques for schizophrenia, making the schizophrenia screening a challenging issue in screening and management of mental diseases. Meanwhile, notice that the study of schizophrenia screening usually involves a vast number of users, and personal privacy information (such as name, address, audio data and basic questionnaires) of these users is collected,

Received September 29, 2022; revised October 31 & December 8, 2022; accepted December 15, 2022.
Communicated by Xiaohong Jiang.
[+] Corresponding author.

stored and transmitted, making the efficient user privacy protection an urgent problem to be solved in such study.

It is notable that for the special mental diseases, early screening and recognition of schizophrenia through voice, behavior and facial expression of users are of particular interest for screening and management of such a mental disease. On one hand, compared with the traditional scale screening method, early screening and identification of schizophrenia through voice, behavior and facial expression of users are more efficient and objective, and can counter users' intentional incompatibility. On the other hand, using user voice and other features for early screening and evaluation of schizophrenia can continuously accumulate large data and feature libraries such as the voice of schizophrenia patients, and promote the use of deep learning, big data analysis and other new technologies to develop new screening methods for mental disorders. More importantly, the usage of the advanced network physical isolation technology can effectively protect the personal privacy of users while conducting early screening and recognition for schizophrenia.

By now, some research efforts have been devoted to the study of schizophrenia screening and user privacy protection, based on biological and behavioral characteristics of users ($e.g.$, facial emotion and vocal expression characteristics). The authors in [1] show that facial emotion defect is one of the most common cognitive disorders, which has been widely studied in various mental diseases, especially schizophrenia. They also summarize facial emotion cognition of schizophrenia patients and investigate some schizophrenic recognition modeling methods. In [2], the authors indicate that verbal and facial emotional information is less consistent in schizophrenia and facial and verbal emotional information is less accurate. In addition, facial encoding studies for schizophrenia have shown a general reduction of facial action responsivity in schizophrenia patients. The authors in [3] demonstrate that voice atypicalities have been a characteristic for schizophrenia since its first definitions, and voice atypicalities may represent a marker of clinical features and social functioning in schizophrenia. The literature [4] focuses on schizophrenic patientn identification and investigate whether the observed changes in speaking behavior and voice sound characteristics were caused by long-term neuroleptic treatment. A set of 12 acoustic variables automatically assessed in a standardized experimental setting allowed an almost perfect discrimination between schizophrenic patients and normal subjects. In [5], the authors show that medical data are in general highly privacy sensitive and thus proper protection on privacy and secure data aggregation/compression are also highly expected in medical data processing. Nevertheless, the existing sensor acquisition technology and data management technology pay less attention to the protection of medical data privacy. The authors in [6] demonstrate that the application of electroencephalogram (EEG) raises privacy concerns, as EEG contains sensitive health and mental information. They propose an augmentation-based source-free adaptation classification approach, which only uses the source model parameters for motor imagery classification instead of the raw EEG data, thus protecting the privacy of the source domain.

It is notable, however, the schizophrenia screening based on machine learning faces dual technical challenges of screening model efficiency and user privacy protection. On one hand, the above available schizophrenia screening methods lack the extraction of fine-grained features of the voice of schizophrenia patients and an effective screening model for schizophrenia. On the other hand, the existing schizophrenia screening lacks data collection and network management technologies with high privacy protection functions. Our results in this work indicate that by exploiting the audio features of the schizophrenia patients and the network physical isolation technology, we can not only provide a full characterization of schizophrenia patients, but also significantly improve the performance of privacy protection. The main contributions of this paper are summarized as follows:

- We first provide extensive experiment results to demonstrate that the audio features of the schizophrenia patients exhibit a good discriminability and stability in screening and recognition of schizophrenia.
- By exploring network physical isolation technology, we design a network security strategy based on wireless sensor network within a private cloud to ensure the absolute security of users' personal privacy.
- By modeling the schizophrenia screening as an SVM-based one-class classification process, we develop a schizophrenia screening framework, which can quickly determine and identify schizophrenic patients, and the determination results can be an important reference for professional doctors and clinical practice.
- Finally, experiment results are provided to illustrate the schizophrenia screening performance of the proposed authentication framework in terms of accuracy. The related screening efficiency issues like the usability to the audio contents, the sensitive to audio features, are also investigated.

2. PROPOSED AUTHENTICATION FRAMEWORK

As shown in Fig. 1, this paper proposes a Support Vector Machine (SVM) based schizophrenia screening and risk assessment framework utilizing network physical isolation within a private cloud for user privacy protection. We first design a network security strategy based on wireless sensor network using network physical isolation technology within a private cloud to ensure the absolute security of users' personal privacy. We further exploit the one-class classification technique to formulate the schizophrenia screening decision process as an SVM model. Finally, we demonstrate the effectiveness of the proposed method through extensive experiments.

2.1 Privacy Protection Based on Network Physical Isolation Within A Private Cloud

In order to ensure the absolute security of users' personal information and voice data during the schizophrenic screening processes, we adopt a physical isolation mechanism to protect personal privacy of users. Physical isolation refers to the technical means of using physical methods to isolate the internal network from the external network to avoid the risk of intrusion or information leakage. Physical isolation is mainly used to solve network security problems, especially in confidential networks that require absolute security. As shown in Fig. 2 when trusted network (private networks for schizophrenia screening) is connected to the Internet, in order to prevent attacks from the Internet and ensure the confidentiality, security, integrity, non-repudiation and high availability of these high-security

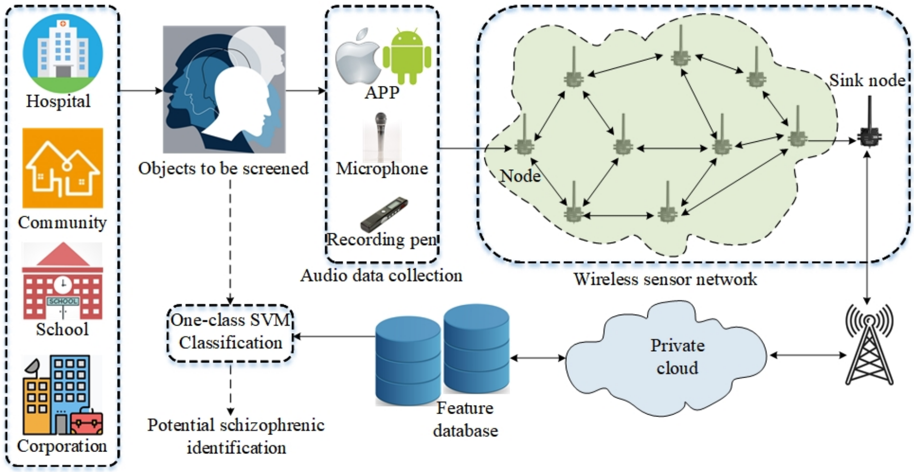


Fig. 1. The processes of the proposed authentication approach for schizophrenic screening (identification).

networks, we use security gatekeeper technology to perform physical isolation of trusted network (intranet) and untrusted network (extranet). The security gatekeeper technology based on physical isolation is a new type of security technology, which can prevent known or unknown attacks against the network layer and operating system layer, and provide a secure, real-time data exchange environment between the internal and external networks while physically separating the internal network from the external network. In addition to

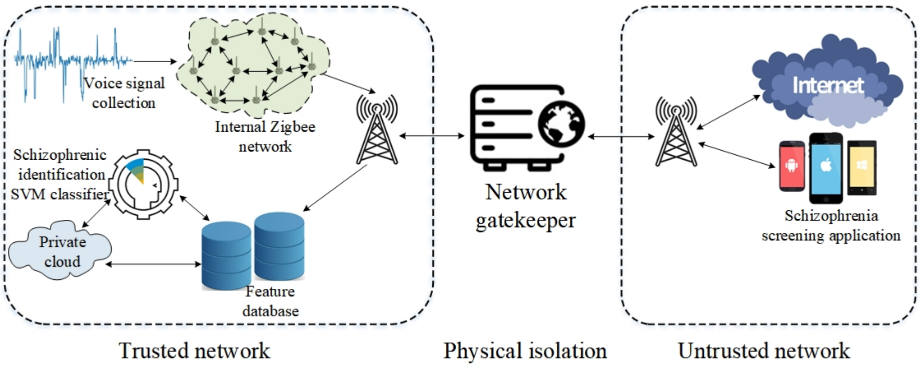


Fig. 2. Privacy protection based on network isolation mechanism.

the physical partition, the security gatekeeper can also allow the secure exchange of data, resources and information between trusted and untrusted networks. As shown in Fig. 2, the biggest benefit of a security gatekeeper is that the physical partition keeps the trusted network secure while allowing online real-time access to the untrusted network. The security gatekeeper connects the two networks at a high speed through the network switch at the physical link level, which physically determines that it is impossible for the user to maintain a stable link between the two networks. In this way, the user's application is still transparent, but the underlying transport has been completely changed. Indeed, it is entirely possible that a hacker compromised an external application server. But, due to the single connectivity of the switch, he/she cannot use the security gatekeeper as a springboard to connect to the service network of the classified network, which fundamentally guarantees the security of the classified network and thus achieves absolute security of user privacy for schizophrenic screening.

2.2 Communication Network Module

In practical applications of our schizophrenic screening framework, we deploy the application of schizophrenic screening in the untrusted network (as shown in Fig. 2), which facilitates the wide-scale application and promotion of the screening model. Therefore, the user audio data to be screened is transmitted to the cloud server center through wireless access communication network and IP-based core network. The IP-based core network is mainly used for long-distance optical fiber transmission and backbone network communication, and usually has a higher communication rate. Generally, wireless communication protocols for access communication networks include ZigBee (IEEE 802.15.4), Bluetooth (IEEE 802.15.1), Wi-Fi (IEEE 802.11) and 3G/4G/ 5G. In the untrusted network of our proposed framework, ZigBee is excluded because most smartphones do not support such a protocol at present. The communication distance between smart devices required for Bluetooth is between 10 meters and 100 meters. Due to insufficient communication rate and communication distance of Bluetooth communication, it is not suitable for transmission of large amount of audio data in schizophrenia screening. Therefore, Bluetooth communication can only be used when it is unavoidable. Wi-Fi is critical for mobile devices interaction with cloud services. Compared with Bluetooth, the communication distance, communication rate and stability of Wi-Fi are far superior to those of Bluetooth. So, the proposed schizophrenic screening framework mainly uses Wi-Fi communication to realize the collection and transmission of audio data from users to be screened.

For some scenarios where Wi-Fi is not available, we adopt cellular network to achieve fast data interaction between mobile terminals and cloud services. Recently, full Wi-Fi coverage has not been achieved, and additional overhead may be incurred when using a cellular network for data transmission. Due to additional overhead, some users may not participate in mobile crowd sensing, especially for non-data-plan users. Therefore, inspired by researches [7], we use the communication model based on the cellular network to resolve the problem of energy consumption and load balancing. As shown in Fig. 3, the specific communication process is as follows. First, we initialize base stations for audio data collection, and allow participants who are data-plan users can apply to become temporary base station nodes. Second, among the base stations within the allowed communication range, the participant chooses the base station with the shortest distance to submit the sensing data ($i.e.$, audio data used for schizophrenic screening). In this case, Bluetooth and Wi-Fi are the most commonly used communication media. Finally, base station nodes forward all collected data to the cloud service center through one or multiple hops. We can see that our communication model can reduce the energy consumption of audio data transmission and balance the load of cellular network while encouraging users to participate in mobile crowd sensing.

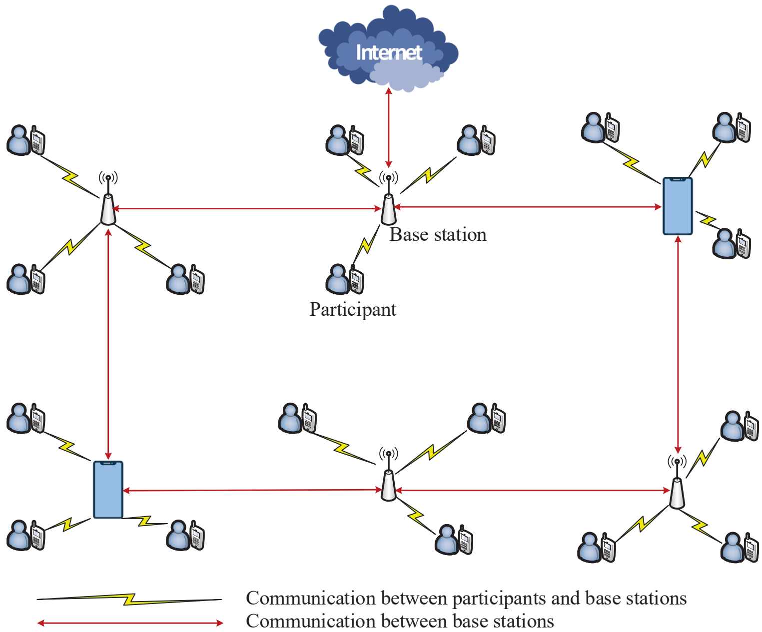


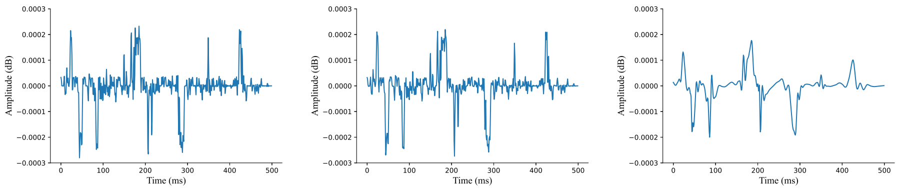
Fig. 3. Communication model in cellular network.

2.3 Raw Data Collection and Preprocessing for Schizophrenic Screening

We develop an application to collect audio data of patients with schizophrenia during rehabilitation in Chuzhou Second People's Hospital, which is a special hospital for mental disorders and has accommodated a large number of schizophrenic patients. Please refer to Section 3.1 for specific audio data acquisition process in our work.

Note that the interference from background noise and invariable magnetic fields is commonly a constant, while user voice signal values are usually time-varying variables. Similar to that in [8], we also adopt a Kalman filtering [9] to estimate the voice signal value. Constant noise from the environment of data collection attached to voice signal component can be reduced and even eliminated through Kalman filtering. Initial raw data directly collected from voice sensors (which are built in the mobile devices) always exists multiple peaks and other interference points due to the non-stationary noise during data collection. The non-stationary noise is in general caused by electromagnetic interference ($e.g.$, changing current and magnetic field, radio frequency signals) and man-made interference ($e.g.$, collision, jitter, and accidental touch on mobile devices). In this paper, we leverage wavelet de-noising to not only reduce the non-stationary noise and interference, but also retain the intrinsic feature of raw sensor data [8, 10].

In Fig. 4 we provide the comparison of initial signal, Kalman filtered signal and wavelet denoised signal. We can see from Fig. 4 that constant noise and non-stationary noise during data collection can be effectively reduced through Kalman filtering and wavelet de-noising.



(a) (b) (c)

Fig. 4. Raw data preprocessing for schizophrenic screening; (a) Initial raw voice data; (b) Kalman filtered voice signal; (c) Wavelet denoised voice signal.

2.4 Feature Extraction

To obtain fine-grained characteristics of audio for each object ($e.g.$, potential schizophrenic) to be screened to accurately identify people with schizophrenia utilizing the user voice data, we extract 9 features from audio data including Short-time Zero-crossing Rate (St-ZCR), Spectral Centroid (SC), Root-mean-Square Energy (RMSE), Sharpness (Sh), Spectral Flatness (SFs), Spectral Flux (SFx), Fundamental Frequency (FF), Loudness (Lo), Mel-scale Frequency Cepstral Coefficients (MFCC). For brevity, we user $F_1 \backsim F_9$ to denote the above 9 audio features, respectively. Specifically, we use F_{Sh}, F_{FF}, F_{Lo} and F_{MFCC} to denote the Sharpness, Fundamental Frequency, Loudness, and Mel-scale Frequency Cepstral Coefficients of the voice signal, and according to [11–16] we can obtain expressions of F_{Sh}, F_{FF}, F_{Lo} and F_{MFCC}, respectively.

St-ZCR is the number of times a frame of speech signal waveform crosses the horizontal axis. Generally, the zero-crossing rate of high-frequency speech is high, and the zero-crossing rate of low-frequency speech is low, so the St-ZCR is an effective parameter to distinguish unvoiced (most energy is concentrated in high frequency) and voiced (most energy is concentrated in low frequency). We use F_{SZ} to denote the St-ZCR of a speech signal, and F_{SZ} is given by

$$F_{\text{SZ}} = \frac{1}{2}\sum_{m=1}^{N} |\text{sgn}(X_n(m)) - \text{sgn}(X_n(m-1))|, \tag{1}$$

where X_n represents the n-th audio data frame signal with N being the length of the frame signal. $sgn(\cdot)$ is the symbolic function, and sgn$(\cdot)$ is written as

$$sgn(x), \begin{cases} = 1, & x > 0, \\ = 0, & x = 0, \\ = -1, & x < 0. \end{cases} \tag{2}$$

We use F_{SC} and F_{RS} to denote the Spectral Centroid and Root-mean-Square Energy of the audio data frame signal X_n, respectively, then F_{SC} and F_{RS} can be given by

$$F_{\text{SC}} = \frac{\sum\limits_{m=1}^{N} (m+1)X_n(m)}{\sum\limits_{m=1}^{N} X_n(m)}, \tag{3}$$

$$F_{\text{RS}} = \frac{\sum\limits_{m=1}^{N}(m+1)X_n(m)}{\sum\limits_{m=1}^{N}X_n(m)}. \tag{4}$$

Spectral flatness is a parameter that quantizes the similarity between signal and noise. The speech spectrum tends to peak in the fundamental frequency and harmonic, while the noise spectrum is relatively flat. The greater the flatness of the signal, the greater the possibility that the signal is noise. Let F_{SFs} denote Spectral flatness of the audio data frame signal X_n, and we have

$$F_{\text{SFs}} = \frac{\sqrt[N]{\prod\limits_{m=1}^{N}X_n(m)}}{\frac{1}{N}\sum\limits_{m=1}^{N}X_n(m)}. \tag{5}$$

To quantize the degree of variation between adjacent frames of an audio signal, we use F_{SFx} to denote the Spectral Flux of the audio data frame signal X_n, and F_{SFx} is written as

$$F_{\text{SFx}} = \sum_{m=1}^{N}\Big(\frac{x_n(m)}{\sum\limits_{j=1}^{N}X_n(j)} - \frac{X_{n-1}(m)}{\sum\limits_{j=1}^{N}X_{n-1}(j)}\Big)^2. \tag{6}$$

2.5 One-class SVM Classification

Similar to the classifier adopted in [17–19], we use one-class SVM algorithm according to Schölkopf to obtain a function denoted by f which takes the value +1 in a "small" region for capturing as many data points in the extracted sample space as possible, while excludes points that are not in the feature space. In particular, let $\mathcal{X}$ be the feature space of voice data from schizophrenic patients, and X be an element of such space. We consider $X = (x_1, x_2, \cdots, x_L)$ with x_l representing the l-th attribute of voice data from schizophrenic patients (samples) and $l = 1, 2, \cdots, L$. Since the data points in feature space $\mathcal{X}$ can't be separated by a straight line, we lift feature space $\mathcal{X}$ to a feature space F, where there can be a straight hyperplane that separates the data points of one class from an other. For simplicity, we use Φ be the feature map, and we have $\Phi = \{\mathcal{X} \to F\}$. Use feature space map Φ, the dot product in the image of Φ can be computed by a kernel function K. In this paper, we use Gaussian Radial Base Function to compute the mapped dot product in Φ, and K is given by

$$K(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right), \tag{7}$$

where σ is a kernel parameter of K and $\|x - y\|$ is the support vector type regularization term for the dissimilarity measure.

Base on the map $\Phi = \{\mathcal{X} \to F\}$, we can map the voice data from schizophrenic patients into the feature space F corresponding to the kernel function K. To perform one-class classification for screening out possible schizophrenic patients, we need to find the function $f(\cdot)$ which takes the value +1 for positive sample points ($e.g.$, schizophrenic patients in $\mathcal{X}$) in a "small" region for capturing as many data points in $\mathcal{X}$ as possible, and –1 for the other sample points ($e.g.$, not belonging to positive sample points). In this paper, we separate the dataset in $\mathcal{X}$ from the origin by solving the following quadratic program:

$$\min_{\omega,\, \xi_i,\, \rho} \frac{1}{2}\|\omega\|^2 + \frac{1}{\nu L}\sum_{i=1}^{L}\xi_i - \rho \tag{8}$$

subject to:

$$\begin{aligned} (\omega \cdot \phi(x_i)) \geq \rho - \xi_i,\ \text{for } i = 1, \cdots, L, \\ \xi_i \geq 0,\ \text{for } i = 1, \cdots, L, \end{aligned} \tag{9}$$

where ξ_i is a nonzero slack variable and is employed to penalize the objective function; $\|\omega\|$ is the support vector type regularization term; ν is trade-off between f containing the number of positive samples and the size of the $\|\omega\|$ value; ω and ρ are two parameters for characterizing the hyperplane $f(\cdot)$.

We adopt Lagrange multipliers to solve the proposed minimization problem in Eqs. (8) and (9) (with quadratic programming). For a new sample (user) $X^u = (x_1^u, x_2^u, \cdots, x_L^u)$ to be screened, we determine if X^u belongs to class +1 or –1 by

$$f(x^u) = \mathrm{sgn}((\omega \cdot \Phi(x_i)) - \rho) = \mathrm{sgn}(\sum_{i=1}^{L}\alpha_i K(x^u, x_i) - \rho), \tag{10}$$

where α are the Lagrange multipliers and every $\alpha_i > 0$ is weighted in the decision function and thus "supports" the machine.

3. EXPERIMENT AND ANALYSIS

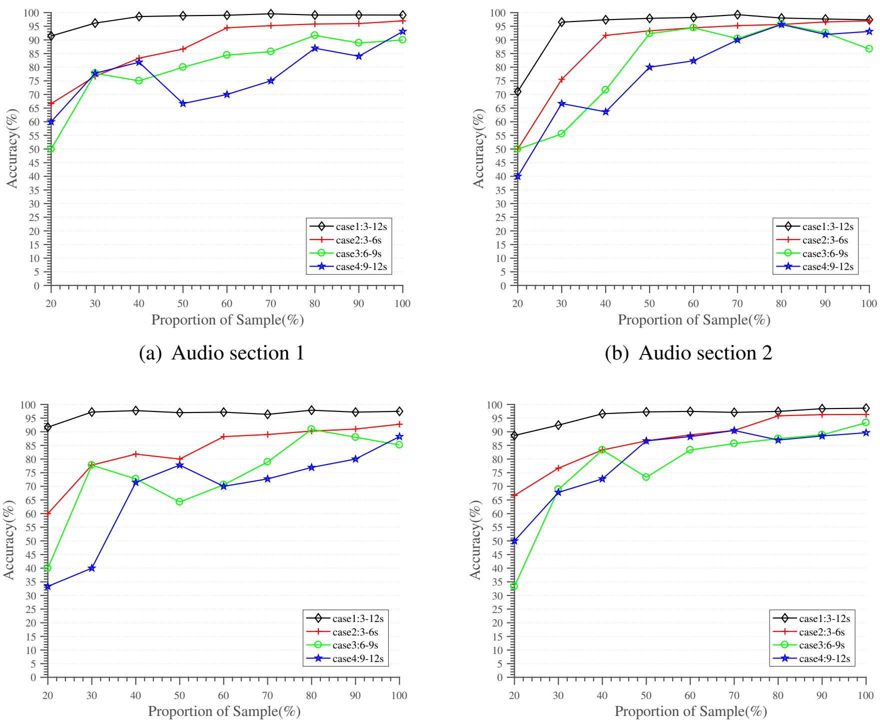
3.1 Performance Metrics and Data Acquisition

In this paper, we adopt classification accuracy as a metric for performance evaluation of the proposed schizophrenia screening method. To evaluate the effectiveness and stability of the proposed schizophrenia screening method in practical applications, we develop an application to collect audio data of patients with schizophrenia during rehabilitation in Chuzhou Second People's Hospital, which is a special hospital for mental disorders and has accommodated a large number of schizophrenic patients. In particular, we select four paragraphs from the Chinese textbooks, and they are from the Chinese Textbooks of 1st grade of elementary school, 3rd grade of elementary school, 6th grade of junior high school and 2nd grade of high school, respectively. For the for paragraphs (audio contents), we control the audio duration of each piece of content to be 15s (calculated according to the speech rate of a normal person). Then we select 179 schizophrenic patients(volunteers) and require them to read the four paragraphs respectively. Finally, we can collect 4 audio data for each schizophrenic patient, namely audio section 1, audio section 2, audio section 3 and audio section 4.

Ideally, For each user we can get 4 audio data segments of length 15s. However, considering that there is a lot of noise in the first 2 seconds of audio data and the data in the last 2 seconds is often blank, we delete the data in the first 2 seconds and the last 2 seconds to ensure the availability of samples. To investigate the effect of audio time on schizophrenia screening, we further sliced the 3-12s audio data as 3-6s, 6-9s and 9-12s.

3.2 Sensitive to Audio Contents

To illustrate the accuracy of schizophrenia screening for the proposed method, we show in Fig. 5 that for audio slices = {case1: 3-12s, case2 :3-6s, case 3: 6-9s, case 4: 9-12s}, how the classification accuracy of the one-class classifier varies with the proportion of training samples. Fig. 5 shows clearly that for all the 4 voice sections, the accuracy for schizophrenia screening using the whole 3-12s (*e.g.*, w/o slice) outperforms the others in terms of accuracy while schizophrenia screening using voice slices 6-9s (case 3) and 9-12s (case 4) obtains worse screening performance. It indicates that utilizing a longer audio for schizophrenia screening usually leads to a more accurate schizophrenia screening effect.



(a) Audio section 1 (b) Audio section 2

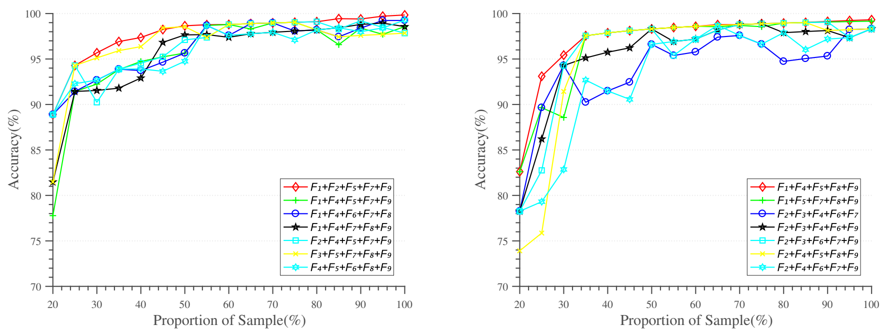
(c) Audio section 3 (d) Audio section 4

Fig. 5. The accuracy of schizophrenia screening for the proposed method; (a) The accuracy of schizophrenia screening using audio section 1; (b) The accuracy of schizophrenia screening using audio section 2; (c) The accuracy of schizophrenia screening using audio section 3; (d) The accuracy of schizophrenia screening using audio section 4.

Specifically, for the 4 audio sections, in case 1 the accuracy of schizophrenia screening is not less than 91.5%, 70.1%, 91.8% and 88.6%, respectively, and the accuracy of schizophrenia screening is always not less than 90% when the proportion of training sample is greater than 40%. However, in case 3 and case 4, the accuracy of schizophrenia screening is almost all lower than 90% for the 4 audio sections, and their accuracy curve fluctuates greatly along with the varying of proportion of training samples. Another observation from Fig. 5 is that for the 4 audio sections, the accuracy of schizophrenia screening in case 2 is very close to case 1, while the required audio data length of case 2 (3s) is much less than that of case 1 (9s). It indicates that when the audio length collecting from the object to be screened is relatively long, we can obtain a significant improvement in the authentication performance in terms of accuracy, but a too large audio length might not be cost efficient since using more audio data in the training and authentication phases will lead to a long screening time without yielding a significant schizophrenia screening performance enhancement. Therefore, it is wise to select a suitable audio length for various schizophrenia screening applications with different authentication performance requirements.

3.3 Sensitive to Audio Features

To evaluate the scalability to the number of features for the proposed schizophrenia screening method, we traverse the combinations of different features ($e.g.$, $F_1 \sim F_9$) for the extracted audio features, and demonstrate that for a relatively small user space size, 5 audio features might be enough to effectively perform schizophrenia screening. We present in Fig. 6 the impact of number of features on schizophrenia screening performance in terms of accuracy with proportion of training samples varying from 20% to 100%. As observed from Fig. 6, the schizophrenia screening in terms of accuracy is greater than 90% and the accuracy is relatively stable under different proportion of training samples in both 3-12s (as shown in Fig. 6(a)) and 3-6s (as shown in Fig. 6(b)) scenarios, thus the proposed schizophrenia screening approach exhibits a good discriminability and stability for lower dimensions features.



(a) Audio length = 3-12s (b) Audio length = 3-6s

Fig. 6. The scalability to the number of features for schizophrenia screening; (a) Different features combinations for audio length = 3-12s; (b) Different features combinations for audio length = 3-6s.

In addition, we can see from Fig. 6(a) and Fig. 6(b) that the accuracy of each combination of 5 features increases as proportion of training sample increases, and there is an optimal combination of 5 features which can obtain optimal performance in terms of accuracy for both audio length equaling to 3-12s and 3-6s. In particular, for the feature combination $[F_1 + F_2 + F_5 + F_7 + F_9]$ in Fig. 6(a) and the feature combination $[F_1 + F_4 + F_5 + F_8 + F_9]$ in Fig. 6(b), as proportion of training sample increases from 20% to 100%, the accuracy increases from 88.8% to 99.8% and from 82.6% to 99.5%, respectively, while the accuracy of other combinations in Fig. 6(a) and Fig. 6(b) the accuracy rate does not increase significantly and there are large fluctuations and noise. Thus, according to different schizophrenia screening requirements, jointly utilizing 5 dimensional feature combination is helpful for maintaining a better authentication accuracy in practical applications of schizophrenia screening.

4. CONCLUDING REMARKS

By exploiting the fine-grained voice features and network isolation privacy protection, we proposed a novel schizophrenia screening and risk assessment method for potential schizophrenic. The proposed approach can effectively assess the risk of patients with schizophrenia during rehabilitation. The screening results, quantitative indicators and risk values can provide important reference for professional doctors and important data support for clinical diagnosis and treatment. We also demonstrated that the new schizophrenia screening method enables a flexible screening performance control to be achieved by adjusting the system parameters like the audio contents and the number of audio features. Thus, the proposed schizophrenia screening method is expected to serve as a good enhancement and complement to the traditional schizophrenia screening solutions.

ACKNOWLEDGMENT

This work was supported in part by the Key Project of Science Research in Universities of Anhui Province of China (No. KJ2021A1066, KJ2021A1067) and the Chuzhou University Project (HX2022116).

REFERENCES

1. Z. Gao, W. Zhao, S. Liu, Z. Liu, C. Yang, and Y. Xu, "Facial emotion recognition in schizophrenia," *Frontiers in Psychiatry*, Vol. 12, 2021, p. 633717.
2. M. K. Mandal, R. Pandey, and A. B. Prasad, "Facial expressions of emotions and schizophrenia: a review," *Schizophrenia Bulletin*, Vol. 24, 1998, pp. 399-412.
3. A. Parola, A. Simonsen, V. Bliksted, and R. Fusaroli, "Voice patterns in schizophrenia: A systematic review and bayesian meta-analysis," *Schizophrenia Research*, Vol. 216, 2020, pp. 24-40.
4. H. Stassen, M. Albers, J. Püschel, C. Scharfetter, M. Tewesmeier, and B. Woggon, "Speaking behavior and voice sound characteristics associated with negative schizophrenia," *Journal of Psychiatric Research*, Vol. 29, 1995, pp. 277-296.

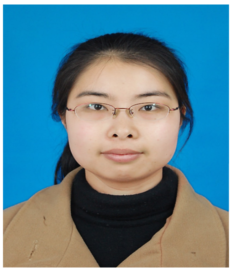
5. L. Wang, L. Li, J. Li, J. Li, B. B. Gupta, and X. Liu, "Compressive sensing of medical images with confidentially homomorphic aggregations," *IEEE Internet of Things Journal*, Vol. 6, 2019, pp. 1402-1409.
6. K. Xia, L. Deng, W. Duch, and D. Wu, "Privacy-preserving domain adaptation for motor imagery-based brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, Vol. 69, 2022, pp. 3365-3376.
7. L. Wang, D. Zhang, Z. Yan, H. Xiong, and B. Xie, "effsense: A novel mobile crowd-sensing framework for energy-efficient and cost-effective data uploading," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 45, 2015, pp. 1549-1563.
8. C. Shen, Y. Li, Y. Chen, X. Guan, and R. A. Maxion, "Performance analysis of multi-motion sensor behavior for active smartphone authentication," *IEEE Transactions on Information Forensics and Security*, Vol. 13, 2018, pp. 48-62.
9. G. Welch, G. Bishop *et al.*, "An introduction to the kalman filter," 1995.
10. G. Zhao, P. Zhang, Y. Shen, and X. Jiang, "Passive user authentication utilizing behavioral biometrics for iiot systems," *IEEE Internet of Things Journal*, Vol. 9, 2022, pp. 12 783-12 798.
11. M. K. R. Hashim, N. A. Ismail, A. A. Saad, P. Y. Chan, U. M. M. Fisol, Z. M. Ripin, H. T. Yow, C. H. Tan, U. H. Hii, and T. S. Saw, "Sound quality analysis for two-way radio under wind noise," in *Proceedings of IEEE International Conference on Smart Instrumentation, Measurement and Applications*, 2013, pp. 1-5.
12. G. Mahé, L. Moisan, and M. Mitrea, "An image-inspired audio sharpness index," in *Proceedings of the 25th European Signal Processing Conference*, 2017, pp. 683-687.
13. C.-T. Do, D. Pastor, and A. Goalic, "On normalized MSE analysis of speech fundamental frequency in the cochlear implant-like spectrally reduced speech," *IEEE Transactions on Biomedical Engineering*, Vol. 57, 2010, pp. 572-577.
14. P. Jain and R. B. Pachori, "Event-based method for instantaneous fundamental frequency estimation from voiced speech based on eigenvalue decomposition of the hankel matrix," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, 2014, pp. 1467-1482.
15. A. Varsavsky and H. J. McDermott, "Application of real-time loudness models can improve speech recognition for cochlear implant users," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 21, 2013, pp. 81-87.
16. C.-T. Do, D. Pastor, and A. Goalic, "On the recognition of cochlear implant-like spectrally reduced speech with MFCC and HMM-based ASR," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, 2010, pp. 1065-1068.
17. B. Liu, Y. Xiao, P. S. Yu, L. Cao, Y. Zhang, and Z. Hao, "Uncertain one-class learning and concept summarization learning on uncertain data streams," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, 2014, pp. 468-484.
18. G. Ratsch, S. Mika, B. Scholkopf, and K.-R. Muller, "Constructing boosting algorithms from svms: an application to one-class classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 2002, pp. 1184-1199.
19. K.-S. Goh, E. Chang, and B. Li, "Using one-class and two-class SVMs for multiclass image annotation," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, 2005, pp. 1333-1346.

Xianchuan Ling graduated from the clinical medicine major of Wannan Medical College in July, 1990. He has been working in the First People's Hospital of Chuzhou and the Second People's Hospital of Chuzhou, specializing in the clinical diagnosis and treatment of psychiatry, and he is now with the title of deputy chief physician. His main research direction is psychiatry, psychological clinical, early screening, diagnosis and treatment of mental disorders.

Xin Zhang graduated from Anhui Normal University with the bachelor degree in Mental Health Education in December, 2012. He got his master degree in Developmental and Educational Psychology from Southwest University in June, 2014. Since July 2021, he has been working for the Second People's Hospital of Chuzhou specializing in disease prevention and control. His major is early screening of mental disorders and psychological counseling.

Huiru Wang graduated from Bengbu Medical College with the bachelor degree in nursing in December 2003. Since 2014, she has been responsible for the nursing management of the psychiatric department involved the First People's Hospital of Chuzhou and the Second People's Hospital of Chuzhou, the management and treatment project of severe mental disorders in the whole city, and related work of disease prevention and control. Her main research interests include management and treatment programs of severe mental disorders (including screening, assessment, health education and rehabilitation of mental disorders) and community rehabilitation management of mental disorders with the chief nurse of psychiatric department.

Guozhu Zhao received his BS degree in Computer Science and Technology from Anhui Normal University, China, in 2008, and MS degree in Computer Application Technology from Changzhou University, China, in 2011. He is an Associate Professor with the School of Computer and Information Engineering at Chuzhou University (Chuzhou, China). He is currently pursuing his Ph.D. degree at the School of Systems Information Science at Future University Hakodate (Hakodate, Japan). His research interests include network security, user authentication, and behavioral biometrics.



Huanhuan Zhao received her Ph.D. degree from the University of Science and Technology of China in 2021. She is a Lecturer at School of Computer and Information Engineering, ChuZhou University, China. Her research interests include wireless communication networks and health information technology.

Shenghui Zhao received her Ph.D. degree from Southeast University of China in 2013. Now her interests include cloud computing, big data and Internet of things. She is a Professor at the School of Computer and Information Engineering and Head of the research group of (Intelligent Information Processing) and big data, Chuzhou University in China.

Zenghui Ding received his Ph.D. degree from the University of Science and Technology of China in 2017, He is an Associate Professor of the Hefei Institute of Physical Science, Chinese Academy of Sciences (Hefei, China). His research interests include brain-inspired computing and substance dependence.